Math 5490 – Homework 3: Due Feb 22 by 11:59pm

Instructions:

- Complete the problems below, and submit your solutions and Python code by uploading them to the Google form: https://forms.gle/29WfqTaEGYi31NiJ7
- Submit all your Python code in a single .py file using the function templates given in each problem. I will import your functions from this file and test your code.
- If you use LaTeX to write up your solutions, upload them as a pdf file. Students who use LaTeX to write up their solutions will receive bonus points on the homework assignment (equivalent to 1/3 of a letter grade bump).
- If you choose to handwrite your solutions and scan them, please either use a real scanner, or use a smartphone app that allows scanning with you smartphone camera. It is not acceptable to submit photos of your solutions, as these can be hard to read.

Problems:

1. Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^n$. Recall that the least squares problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|^2$$

has a unique solution $\mathbf{x}^*$ belonging to img A^T (this is the minimal norm solution). Show that when AA^T is invertible we can express $\mathbf{x}^*$ as

$$\mathbf{x}^* = A^T(AA^T)^{-1}b.$$

2. Complete the following parts.
   (i) Show that the linear function $F(\mathbf{x}) = \mathbf{x}\cdot\mathbf{w} + b$ is a convex function.
   (ii) Show that the Euclidean norm $F(\mathbf{x}) = \|\mathbf{x}\|$ is a convex function. [Hint: Use the triangle inequality and the definition of convexity.]
   (iii) Let $F : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R} \to \mathbb{R}$ be convex functions. Assume G is non-decreasing (i.e., $G(x) \le G(y)$ whenever $x \le y$). Show that the composition $G \circ F$, defined by

$$(G \circ F)(x) = G(F(x))$$

   is a convex function. [Hint: Use the definition of a convex function directly.]

3. For $\beta > 0$ define

$$\psi_\beta(x) = \frac{1}{\beta}\log(1 + e^{\beta x}).$$

   (i) Show that

$$\lim_{\beta\to\infty} \psi_\beta(x) = x_+ := \max\{x, 0\}.$$

   (ii) Show that

$$\psi_\beta'(x) = \frac{1}{1 + e^{-\beta x}}.$$

(iii) Show that

$$\lim_{\beta\to\infty} \psi'_\beta(x) = \begin{cases} 1, & \text{if } x > 0, \\ \frac{1}{2}, & \text{if } x = 0, \\ 0, & \text{if } x < 0. \end{cases}$$

This problem shows that the positive part x_+ can be well-approximated by the smooth function ψ_β.

4. In light of problem 3, we consider the smooth approximation of the soft-margin SVM problem given by

$$\min_{\mathbf{w},b} E(\mathbf{w}, b),$$

where $\mathbf{w} \in \mathbb{R}^n$, $b \in \mathbb{R}$ and

$$E(\mathbf{w}, b) = \lambda\|\mathbf{w}\|^2 + \frac{1}{m}\sum_{i=1}^{m} \psi_\beta(1 - y_i(\mathbf{x}_i \cdot \mathbf{w} - b)). \tag{1}$$

(i) Explain why E is a *convex* function of $\mathbf{w}$ and b [Hint: Use problem 2].

(ii) Show that

$$\nabla_{\mathbf{w}} E(\mathbf{w}, b) = 2\lambda\mathbf{w} - \frac{1}{m}\sum_{i=1}^{m} \frac{y_i \mathbf{x}_i}{1 + e^{-\beta(1 - y_i(\mathbf{x}_i \cdot \mathbf{w} - b))}},$$

and

$$\nabla_b E(\mathbf{w}, b) = \frac{1}{m}\sum_{i=1}^{m} \frac{y_i}{1 + e^{-\beta(1 - y_i(\mathbf{x}_i \cdot \mathbf{w} - b))}}.$$

(iii) Write a Python program to train an SVM by minimizing E via gradient descent

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \alpha\nabla_{\mathbf{w}} E(\mathbf{w}_k, b_k)$$
$$b_{k+1} = b_k - \alpha\nabla_b E(\mathbf{w}_k, b_k).$$

Test your program at first on some synthetic data, like the two-point example from class and the course textbook (where $\mathbf{x}_1 = \mathbf{z}$ and $\mathbf{x}_2 = -\mathbf{z}$). Then test your algorithm on pairs of MNIST digits. Try different pairs of MNIST digits. Which are easiest to separate? You can use this notebook as a starting place.